Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming

Duo Chan[1], Elizabeth C. Kent[2], David I. Berry[2] & Peter Huybers[1]*

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA.

[2]National Oceanography Centre, Southampton, UK.

*e-mail: phuybers@fas.harvard.edu

Existing estimates of sea surface temperatures (SSTs) indicate that, during the early twentieth century, the North Atlantic and Northeast Pacific oceans warmed by twice the global average, whereas the Northwest Pacific Ocean cooled by a magnitude equal to the global average[1–4]. Such a heterogeneous pattern suggests first-order contributions from regional variations in forcing or in ocean–atmosphere heat fluxes[5,6]. These older SST estimates are, however, derived from measurements of water temperatures in ship-board buckets, and must be corrected for substantial biases[7–9]. Here we show that correcting for offsets among groups of bucket measurements leads to SST variations that better correlate with nearby land temperatures and are more homogeneous in their pattern of warming. Offsets are identified by systematically comparing nearby proximal SST observations among different groups[10]. Correcting for offsets in German measurements decreases warming rates in the North Atlantic, whereas correcting for Japanese offsets leads to increased and more uniform warming in the North Pacific. Japanese offsets in the 1930s primarily result from records having been truncated to whole-degrees Celsius when digitized in the 1960s. These findings underscore the fact that historical SST records reflect both physical and social dimensions, and suggest that further opportunities exist for improving the accuracy of historical SST records[9,11].

According to recent estimates from the National Oceanic and Atmospheric Administration (NOAA)[1], global average SST warmed by 0.43 °C between 1908 and 1941. Whereas the North Atlantic warmed by 0.82 °C, the North Pacific showed a bimodal structure,

with the Northwest Pacific cooling by −0.39 °C and the Northeast Pacific warming by 1.02 °C. Other gridded SST products give similarly disparate SST trends for the early twentieth century (Table 1 and Extended Data Fig. 1), and together these SST estimates suggest that internal modes of variability strongly contributed to early-twentieth-century climate change. Specifically, the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation have been suggested to account for regional variations as well as more than half of the global warming between 1908 and 1941 (refs [6,12]). Model simulations of the atmospheric and oceanic response to prescribed radiative forcing do not, however, reproduce either the magnitude[13,14] or the pattern[5,15] of the early-twentieth-century warming seen in observations (Extended Data Fig. 1e). Difficulty in reproducing observations has been suggested to arise from deficiencies in how radiative forcing is prescribed[16] or from model limitations in representing internal climate variability[17,18].

Another possibility is that observational estimates of SST changes contain undetected biases, for which there are some precedents. Difficulty in simulating a slowdown in global warming between 1997 and 2012 was partly reconciled by revising SST estimates[19], amongst other considerations[20]. In another study[21], a jump in global temperature by 0.3 °C in 1945 was attributed to offsets between engine-room intake and bucket SST estimates.

The four major SST products covering the early twentieth century each rely upon the International Comprehensive Ocean-Atmosphere Data Set (ICOADS)[22], whose latest release is 3.0. It is estimated that 94% of observations between 1908 and 1941 were from buckets (Fig. 1). Bucket measurements of SST are biased by evaporative, sensible and solar heat fluxes that depend on a range of factors, including weather, ship deck height and bucket type[7]. For example, a canvas bucket left on deck for three minutes under typical wind and other weather conditions can give water temperatures that are approximately 0.5 °C cooler than a wooden bucket measured using the same protocol[7,9].

Foregoing corrections for bucket-measurement biases have involved assumptions that these biases change smoothly in space or time[1,4]. HadSST3, for example, represents transitions between wooden buckets and less-insulated canvas buckets by using globally uniform, linear

weights[4]. Another method assumes that differences between SST and night-time marine air temperatures remain similar to an average for 1971–2000 (ref. [1]). Simple assumptions regarding bucket corrections are commensurate with the limited metadata available for determining observational characteristics[2,4,7], but are questionable on the basis that ICOADS contains observations from a wide variety of ships that sailed for different purposes and sampled the ocean unevenly[22].

More comprehensive bucket corrections are made possible by calculating offsets between groups of SST measurements that are nearby in space and time[10]. Specifically, bucket SST measurements that come from distinct groups of ships within 300 km and 2 days of one another can be differenced, giving a data set of 6.1×10^6 SST comparisons between 1908 and 1941. Groups are designated according to nations and 'decks', with the latter term inherited from the fact that marine observations were stored using decks of punch cards. SST differences are analysed using a linear-mixed-effects (LME) methodology after accounting for climatological effects associated with location, day of year, and hour of day[10]. Mean offsets between groups of SST measurements range from $-0.3\ °C$ to $+0.6\ °C$ (Supplementary Table 1). Of the 46 nation–deck groups that contribute SST observations between 1908 and 1941, 21 have significant offsets ($P < 0.05$), and 6 remain significant after applying a Bonferroni correction[23] for multiple hypothesis testing ($P < 0.05/n$, $n = 46$; Fig. 2).

The presence of systematic offsets between groups of measurements, combined with changes in the distribution of these groups over time (Fig. 1), is liable to introduce spurious SST trends. To diagnose these trends, we first construct a bucket SST data set that is corrected for biases common to all groups of bucket SSTs following the same approach used for HadSST3 (refs [4,7]). This reference data set, ICOADSa, is then further corrected for offsets between groups of ships to obtain our best estimate of SST trends, ICOADSb (see Methods). Comparing the trends in ICOADSa (Fig. 3a) against the difference in trends between ICOADSa and ICOADSb between 1908 and 1941 (Fig. 3b) shows a spatial anticorrelation (Pearson's r-value) of -0.50 ± 0.03. All uncertainties are reported with two standard deviations (s.d.) unless otherwise noted. Similar anticorrelations between our groupwise trend corrections

and baseline SST trends are found for each of four major SST products (Table 1), indicating that some of the structure in early-twentieth-century SST trends reflects offsets between nations and decks.

The corrections included in ICOADSb result in a more homogeneous warming pattern (Extended Data Fig. 1f). North Pacific trends change from 0.31 ± 0.03 °C per 34 years in ICOADSa to 0.55 ± 0.11 °C per 34 years in ICOADSb, and North Atlantic trends from 0.85 ± 0.03 °C per 34 years to 0.65 ± 0.11 °C per 34 years (Fig. 4). Results are also temporally more homogeneous, with ICOADSa and other SST estimates indicating a slight cooling between 1920 and 1941 in the North Pacific, but ICOADSb showing a continuous warming trend throughout the early twentieth century (Fig. 4a). Furthermore, whereas ICOADSa indicates that 6% of the 5° grid boxes for which SST trends are computed contain significant cooling ($P < 0.05$), ICOADSb indicates that only 2% of grid boxes contain significant cooling (Fig. 3). Our focus is on the 1908–1941 interval for consistency with a previous study[12], but results are similar if neighbouring starting and ending years are chosen instead (Supplementary Table 2).

The prominent revision to Pacific SST trends relates to a change in offsets identified in the Japanese Kobe Collection, where a mean offset of 0.07 ± 0.12 °C between 1908 and 1930 drops to -0.28 ± 0.13 °C between 1935 and 1941 in the North Pacific (Extended Data Fig. 2c). One indication of why offsets change comes from considering the distribution of trailing digits in the Kobe Collection records, which are distributed across decks 118, 119 and 762. Before 1932, Kobe Collection records come from deck 762, and the distribution of trailing digits in reported SSTs is consistent with measurements that were originally recorded in whole-degrees Fahrenheit, then converted to Celsius, and finally rounded to tenths-of-a-degree (Extended Data Fig. 3a). Such unit conversion and rounding increases noise but results in negligible systematic offsets[24]. Between 1933 and 1941, however, more than 99% of Kobe Collection records come from deck 118 and are archived in whole-degrees Celsius. Examination of the reference manual for deck 118 reveals that digitization of these observations involved truncation[25] (Extended Data Fig. 3d).

If, before truncation, deck 118 contained the same distribution of trailing digits as found in deck 762, truncation would be expected to introduce a cold offset of −0.46 °C. We find an offset of −0.35 ± 0.07°C, with the smaller magnitude possibly reflecting the presence of additional offsets between decks. Cold offsets identified in the Kobe Collection during the 1930s are robust to whether or not decks 762 and 118 are treated as distinct groups in our analysis, because these decks are well separated in time and our LME model allows offsets to vary temporally (see Extended Data Figs. 2d and 4 for a sensitivity analysis). Kobe Collection deck 119 is also truncated and found to have a cold offset, but there is little consequence for early-twentieth-century trends because deck 119 spans 1951 to 1961. The biases identified here are associated with data-management practices and are not addressed in physical models used for correcting bucket temperatures. Furthermore, truncation biases in SSTs could, in principle, be identified by comparing against night-time marine air temperatures, but in the case of decks 118 and 119 this other indicator of surface temperature was also truncated.

North Atlantic warming in ICOADSb is revised downwards by −0.20 ± 0.11 °C between 1908 and 1941, primarily as a consequence of the revision of SSTs from deck 156 and German deck 192 (Fig. 4b and Extended Data Fig. 2b). Deck 156, a group without country information, is biased relatively cold by −0.22 ± 0.12 °C during 1908–1941 in the North Atlantic (Extended Data Fig. 2c). Deck 156 contributes 41% of North Atlantic observations between 1908 and 1912, but this contribution drops to less than 1% during World War I and returns only to 7% thereafter. Deck 156 corrections predominantly warm SST estimates earlier in the twentieth century, decreasing the warming trend by −0.05 ± 0.03 °C per 34 years. Another major correction involves making German deck 192 cooler by 0.33 ± 0.13 °C between 1920 and 1941 (Extended Data Fig. 2c), thereby decreasing estimates of North Atlantic warming by −0.08 ± 0.03 °C per 34 years. A possible clue to the origins of this offset is that German deck 192 shows an increasing portion of SSTs that are reported in whole-degrees Celsius after the 1920s (Extended Data Fig. 3c), although no major modification of the data is indicated by available documentation.

A test of whether our groupwise SST corrections are skilful is made by comparing ICOADSa and ICOADSb against land-based air-temperature records. We compare SSTs from the Northwest Pacific and Northwest Atlantic against adjacent land temperatures from CRUTEM4 (ref. [26]), because these regions show large corrections near a network of land-temperature stations. The Pearson cross-correlation for interannual temperature variability in the Northwest Pacific between 1908 and 1941 increases from 0.67 when using ICOADSa to 0.85 when using ICOADSb. Furthermore, the 1908–1941 trend increases from -0.06 ± 0.04 °C per 34 years in ICOADSa to 0.34 ± 0.11 °C per 34 years in ICOADSb, bringing it into agreement with the 0.33 °C per 34 years trend in CRUTEM4 (Table 1 and Extended Data Fig. 5). Improved agreement for the Northwest Atlantic is more modest. Northwest Atlantic correlation increases from 0.65 to 0.71 from ICOADSa to ICOADSb, and trends decrease from 0.91 ± 0.05 °C per 34 years to 0.64 ± 0.13 °C per 34 years, but this is still significantly higher than the CRUTEM4 trend of 0.32 °C per 34 years.

The Pacific Decadal Oscillation[18] (PDO) shifts towards an increasingly positive phase over 1908–1941 (ref. [27]). The magnitude of the PDO trend is smaller in ICOADSb than in ICOADSa (Table 1), but both are consistent with trends found in North Pacific sea level pressure fields[28] (Extended Data Fig. 6).

Equally important to changed SST trends is the greater uncertainty estimated to accompany these trends (Extended Data Fig. 7). The global average SST trend in HadSST3 between 1908 and 1941 has a reported 2 s.d. uncertainty of 0.03 °C per 34 years, whereas accounting for groupwise offsets in ICOADSb reveals an uncertainty of 0.10 °C per 34 years. Larger uncertainties in ICOADSb reflect the fact that averaging repeated observations within a group does not decrease systematic groupwise errors. One implication is that the expected correspondence between observed and simulated trends should be revised downwards. These results have implications for attribution of early-twentieth-century warming and extreme events[5,27,29]. Another implication is that the scope for further improvement of regional temperature estimates associated with better diagnosing and correcting for groupwise biases[9] is greater than previously recognized.

Finally, we briefly explore the implications of our results for model–data mismatches during the early twentieth century. Differences in rates of warming in the North Atlantic and North Pacific reduce from 0.54 ± 0.03 °C per 34 years in ICOADSa to 0.10 ± 0.07 °C per 34 years in ICOADSb. These revised interbasin trend differences are consistent with that of 0.00 ± 0.40 °C per 34 years found in the early-twentieth-century simulations from the Fifth Climate Model Intercomparison Project[30] (CMIP5; Extended Data Fig. 1e). But we note that the global-average rate of SST warming in ICOADSb is 0.56 ± 0.10 °C per 34 years, and that the same domain in the CMIP5 ensemble warms by only 0.19 ± 0.17 °C per 34 years—a discrepancy in warming rates noted previously for other SST estimates[27]. The model–data mismatch in rates of overall warming highlights the importance of continuing to investigate forcing, sensitivity and internal variability within model simulations along with corrections to historical SST estimates.

<jrn>1. Huang, B. et al. Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J. Clim.* **30**, 8179–8205 (2017).</jrn>

<jrn>2. Hirahara, S., Ishii, M. & Fukuda, Y. Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **30**, 57–75 (2014).</jrn>

<conf>3. Kennedy, J., Rayner, N., Saunby, M. & Millington, S. Bringing together measurements of sea surface temperature made in situ with retrievals from satellite instruments to create a globally complete analysis for 1850 onwards, HadISST2. In *EGU General Assembly Conference Abstracts* (2013).</conf>

<jrn>4. Kennedy, J., Rayner, N., Smith, R., Parker, D. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res. D* **116**, D14104 (2011).</jrn>

<jrn>5. Delworth, T. L. & Knutson, T. R. Simulation of early 20th century global warming. *Science* **287**, 2246–2250 (2000).</jrn>

<jrn>6. Tung, K.-K. & Zhou, J. Using data to attribute episodes of warming and cooling in instrumental records. *Proc. Natl Acad. Sci. USA* **110**, 2058–2063 (2013).</jrn>

<jrn>7. Folland, C. & Parker, D. Correction of instrumental biases in historical sea surface temperature data. *Q. J. R. Meteorol. Soc.* **121**, 319–367 (1995).</jrn>

<jrn>8. Smith, T. M. & Reynolds, R. W. Bias corrections for historical sea surface temperatures based on marine air temperatures. *J. Clim.* **15**, 73–87 (2002).</jrn>

<jrn>9. Kent, E. C. et al. A call for new approaches to quantifying biases in observations of sea surface temperature. *Bull. Am. Meteorol. Soc.* **98**, 1601–1616 (2017).</jrn>

<jrn>10. Chan, D. & Huybers, P. Systematic differences in bucket sea surface temperature measurements amongst nations identified using a linear-mixed-effect method. *J. Clim.* https://doi.org/10.1175/JCLI-D-18-0562.1 (2019).</jrn>

<jrn>11. Davis, L. L., Thompson, D. W., Kennedy, J. J. & Kent, E. C. The importance of unresolved biases in 20th century sea-surface temperature observations. *Bull. Am. Meteorol. Soc.* 621–629 (2018).</jrn>

<jrn>12. Tokinaga, H., Xie, S.-P. & Mukougawa, H. Early 20th-century Arctic warming intensified by Pacific and Atlantic multidecadal variability. *Proc. Natl Acad. Sci. USA* **114**, 6227–6232 (2017).</jrn>

<jrn>13. Crook, J. A. & Forster, P. M. A balance between radiative forcing and climate feedback in the modeled 20th century temperature response. *J. Geophys. Res. D Atmospheres* **116,** D17108 (2011).</jrn>

<jrn>14. Knutson, T. R., Zeng, F. & Wittenberg, A. T. Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth-century simulations. *J. Clim.* **26**, 8709–8743 (2013).</jrn>

<jrn>15. Stott, P. A. et al. External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290**, 2133–2137 (2000).</jrn>

<jrn>16. Myhre, G. et al. Anthropogenic and natural radiative forcing. *Clim. Change* **423**, 658–740 (2013).</jrn>

<jrn>17. Schlesinger, M. E. & Ramankutty, N. An oscillation in the global climate system of period 65–70 years. *Nature* **367**, 723–726 (1994).</jrn>

<jrn>18. Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).</jrn>

<jrn>19. Karl, T. R. et al. Possible artifacts of data biases in the recent global surface warming hiatus. *Science* **348**, 1469–1472 (2015).</jrn>

<jrn>20. Medhaug, I., Stolpe, M. B., Fischer, E. M. & Knutti, R. Reconciling controversies about the 'global warming hiatus'. *Nature* **545**, 41–47 (2017).</jrn>

<jrn>21. Thompson, D. W., Kennedy, J. J., Wallace, J. M. & Jones, P. D. A large discontinuity in the mid-Twentieth Century in observed global-mean surface temperature. *Nature* **453**, 646–649 (2008).</jrn>

<jrn>22. Freeman, E. et al. ICOADS release 3.0: a major update to the historical marine climate record. *Int. J. Climatol.* **37**, 2211–2232 (2017).</jrn>

<jrn>23. Bonferroni, C. Teoria statistica delle classi e calcolo delle probabilita. *Pubbl. R Ist. Sup. Sci. Econ. Commer. Fir.* **8**, 3–62 (1936).</jrn>

<jrn>24. Rhines, A., Tingley, M. P., McKinnon, K. A. & Huybers, P. Decoding the precision of historical temperature observations. *Q. J. R. Meteorol. Soc.* **141**, 2923–2933 (2015).</jrn>

<jrn>25. Wilkinson, C. et al. Recovery of logbooks and international marine data: the RECLAIM project. *Int. J. Climatol.* **31**, 968–979 (2011).</jrn>

<jrn>26.   Jones, P. et al. Hemispheric and large-scale land-surface air temperature variations: an extensive revision and an update to 2010. *J. Geophys. Res. D* **117**, D05127 (2012).</jrn>

<jrn>27.   Hegerl, G. C., Brönnimann, S., Schurer, A. & Cowan, T. The early 20th century warming: anomalies, causes, and consequences. *Wiley Interdiscip. Rev. Clim. Change* **9**, e522 (2018).</jrn>

<jrn>28.   Compo, G. P. et al. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).</jrn>

<jrn>29.   Seager, R. et al. Would advance knowledge of 1930s SSTs have allowed prediction of the Dust Bowl drought? *J. Clim.* **21**, 3261–3281 (2008).</jrn>

<jrn>30.   Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).</jrn>

Author contributions D.C. and P.H. conceived and designed the study; D.C. performed the analyses; and all authors contributed to interpreting results and writing the paper.

Competing interests: The authors declare no competing interests.

Additional information

Extended data is available for this paper at [DOI link].

Supplementary information is available for this paper at [DOI link].

Reprints and permissions information is available at www.nature.com/reprints.

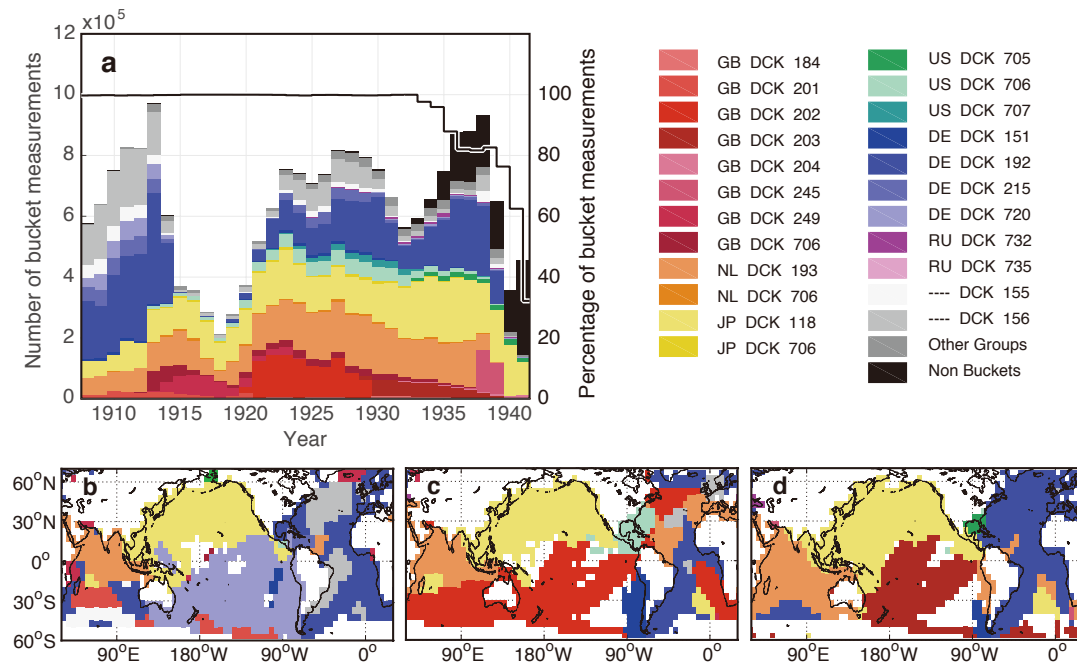**Correspondence and requests for materials** should be addressed to P.H.

Fig. 1 | Groups of bucket SST measurements. a, Left-hand *y*-axis: number of bucket SST

measurements from individual groups identified by country and deck information in

ICOADS3.0. Country name abbreviations are: DE, Germany; GB, Great Britain; JP, Japan;

NL, The Netherlands; RU, Russia; US, United States; and —, missing. Groups having fewer

than 100,000 measurements are labelled as 'other groups'. Decks 118 and 762 are combined

into 'JP DCK 118' because they are both Japanese Kobe Collection decks. Right-hand *y*-axis:

percentage of measurements that have come from buckets, showing that nearly all observations

before 1935 are from buckets (black line). **b–d**, Maps indicating nations that contribute the

most observations within 5° × 5° grids for the periods 1908–1918 (**b**), 1919–1928 (**c**) and

1929–1941 (**d**). White grid boxes have fewer than three years of data.
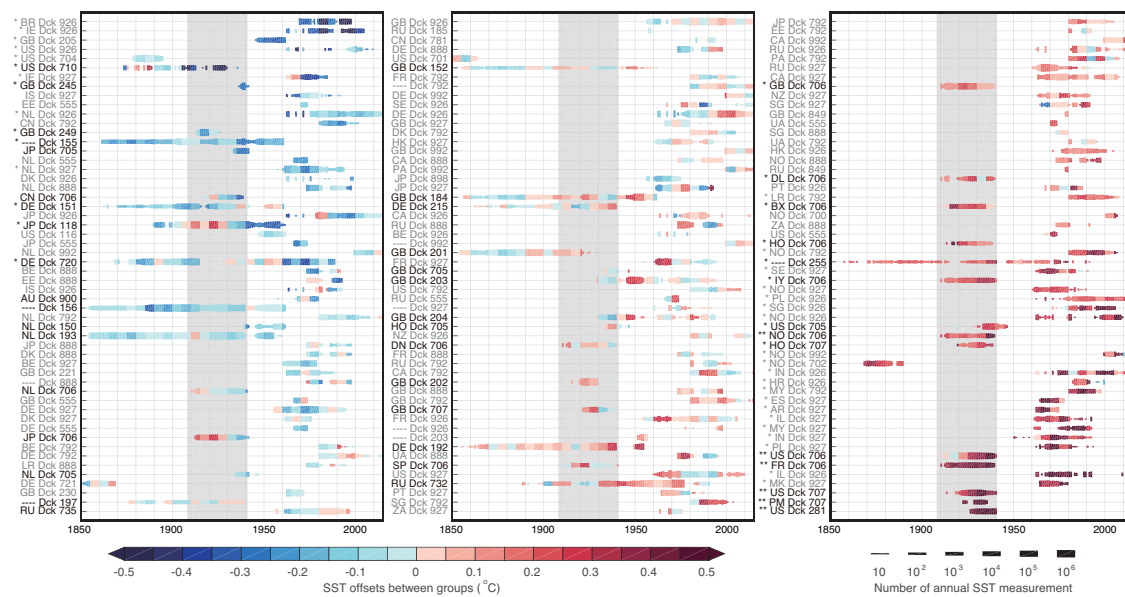
Fig. 2 | Relative offsets between groups of bucket SSTs. Groups for which fixed effects differ significantly from zero are indicated by * ($P < 0.05$). The 46 out of 162 groups that contribute data between 1908 and 1941 are indicated in black, and those remaining significant after a Bonferroni correction ($P < 0.05/46$) are indicated by **. Shading indicates the sum of fixed and five-yearly effects (regional effects are not shown). Bar widths indicate the number of SST measurements contributed by each group for each year. Abbreviated country names correspond to those in Fig. 1 and are listed in Supplementary Table 1.
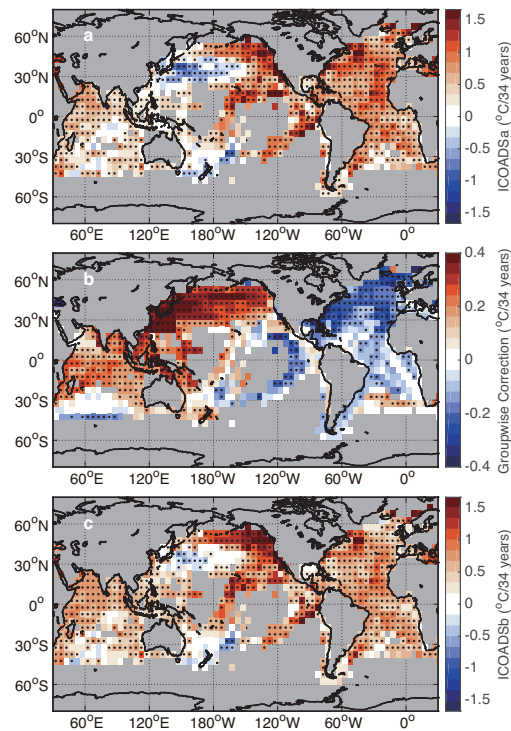
Fig. 3 | Maps of SST trends and corrections between 1908 and 1941. a, SST trends in ICOADSa are similar to patterns found in existing SST estimates (Extended Data Fig. 1). **b**, Trends associated with the corrections for groupwise offsets. Note that panel **b** is plotted on a different colour scale. **c**, SST trends in ICOADSb after applying groupwise corrections. Areas in grey are inadequately sampled for purposes of calculating trends (see Methods). Dots indicate significant trends ($P < 0.05$). In ICOADSb (**c**), 77% of boxes show statistically significant warming, whereas only 2% show significant cooling. By contrast, in ICOADSa (**a**), 6% of boxes indicate significant cooling.
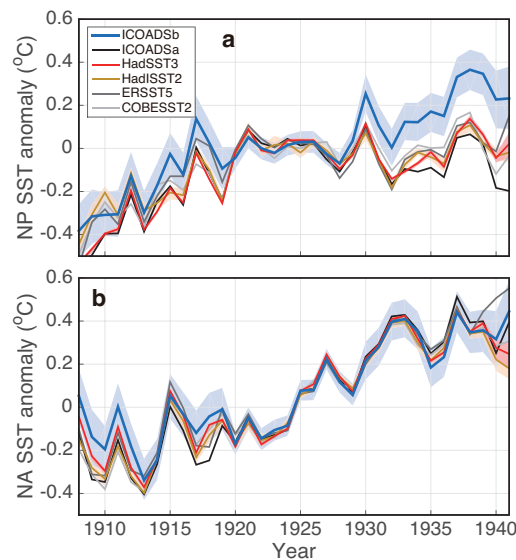
Fig. 4 | Diverging estimates of regional temperature variations. a, **b**, Annual SST anomalies from different datasets in the North Pacific (**a**) and North Atlantic (**b**) oceans. Anomalies are relative to the 1920–1929 mean of each SST estimate. ICOADSb shows greater warming in the North Pacific and less warming in the North Atlantic relative to previous estimates. Uncertainties associated with ICOADSb (blue shading, 2 s.d.) are for annual average SSTs for each sub-basin, and are an order of magnitude larger than those reported for HadSST3 (red shading). Note that those uncertainties included in HadSST3 are mostly removed when computing the anomaly. The discrepancy in annual average SST uncertainties is larger than the discrepancy for trends (Table 1).

Table 1 | Early-twentieth-century SST trends

	ICOADSa	ICOADSb	ERSST5	COBESST2	HadISST2	HadSST3
Global trend	0.51 ± 0.03	0.56 ± 0.10	0.43	0.48	0.44 ± 0.04	0.47 ± 0.03
N Atlantic trend	0.85 ± 0.03	0.66 ± 0.11	0.82	0.79	0.74 ± 0.04	0.71 ± 0.03
N Pacific trend	0.31 ± 0.03	0.56 ± 0.11	0.37	0.39	0.32 ± 0.04	0.37 ± 0.03
NW Pacific trend	−0.35 ± 0.04	−0.02 ± 0.11	−0.39	−0.14	−0.34 ± 0.06	−0.30 ± 0.04
NE Pacific trend	0.86 ± 0.04	1.03 ± 0.12	1.02	0.85	0.85 ± 0.06	0.94 ± 0.04
East Asia trend	−0.06 ± 0.04	0.34 ± 0.11	−0.31	0.04	−0.09 ± 0.04	0.00 ± 0.04
Eastern US trend	0.92 ± 0.05	0.65 ± 0.13	0.75	0.82	0.61 ± 0.05	0.56 ± 0.05
PDO trend	1.82 ± 0.19	1.53 ± 0.20	2.03 ± 0.15	1.64 ± 0.11	2.02 ± 0.12	1.98 ± 0.18
East Asia air-temperature correlation	0.67	0.85	0.53	0.68	0.69	0.71
Eastern US air-temperature correlation	0.65	0.70	0.67	0.56	0.72	0.76
Correction pattern correlation	−0.49	−0.10	−0.49	−0.43	−0.45	−0.42

Trends are averaged over non-grey areas as in Fig. 3a, with the North Atlantic and North Pacific defined as poleward of 20° N; the Northwest Pacific between 120° E and 180° E, and 25° N and 45° N; and the Northeast Pacific between 120° W and 160° W, and 20° N and 60° N. East Asia and Eastern US regions are shown in Extended Data Fig. 5c. All trends are between 1908 and 1941, and are in units of °C per 34 years, with uncertainties reported at the 2 s.d. level. Each reported SST trend uncertainty includes contributions from bucket corrections, but only ICOADSb also accounts for groupwise offsets. Cross-correlations (Pearson's *r*) are reported between interannual air temperatures[26] and SSTs from East Asia and the Eastern US between 1908 and 1941, and between the spatial patterns of trends in SSTs (Fig. 3a and Extended Data Fig. 1a–d) and groupwise corrections (Fig. 3b).

METHODS

Identification of bucket measurements

To identify bucket measurements, we follow the same procedure used for HadSST3 (ref. [4]), using World Meteorological Organization Report Number 47 (WMO47) and ICOADS metadata. Prior to 1941, all SST measurements are assumed to be from buckets unless explicitly recorded otherwise. Analysis of the amplitude of the diurnal cycle in SST before 1941 supports unidentified records as being overwhelmingly from buckets[31]. From 1941 onwards, if the method of measurement is missing in both WMO47 and ICOADS metadata,

SST measurements are assumed to come from buckets if the associated nations are reported to have at least 95% of their ships making bucket measurements in WMO47.

The nationality corresponding to each bucket SST measurement is determined using ICOADS country-code information and WMO47. If these metadata are unavailable, the first three letters of the ICOADS identification code are matched with international call signs[32], or the first two letters for decks 705, 706, and 707 (ref. [33]). Decks that may include substantial amounts of engine-room-intake measurements (decks 740, 780 and 874) are discarded, but this has little influence because these decks together contribute only 249 measurements between 1908 and 1941 out of a total of 20.5 million measurements.

Bucket corrections

The same methodology used for HadSST3 (ref. [4]) is applied to ICOADSa and ICOADSb to correct for biases common to all groups (see Extended Data Fig. 8 for comparison and Supplementary Table 3 for details). Corrections common to all groups are made using wooden- and canvas-bucket models[7] run at 5° × 5° for individual climatological months. Bucket models are driven by the 1973–2002 monthly climatology of SST, 10-m air temperature, wind speed and specific humidity from the National Oceanography Centre (NOC) version 2.0 surface flux and meteorological data set[34] and an insolation climatology from ERA-interim reanalysis[35].

Additional corrections are applied to groups of bucket observations in ICOADSb that are determined using an LME model:

$$\delta T = X\alpha + Z_y\beta_y + Z_r\beta_r + \beta_\sigma \qquad (1)$$

The vector of temperature differences, δT, is determined from proximal pairs of bucket SST observations that come from ships within 300 km and 2 days of one another that are associated with different nationalities and deck assignments. All bucket SST data identified in ICOADS3.0 between 1850 and 2014 are analysed, yielding 17.8 million paired SST differences (Extended Data Fig. 9). The 1908–1941 period contains a subset of 6.1 million SST differences. SST differences are adjusted for climatological effects associated with location, day of year, and hour of day[10].

SST differences contained in $\boldsymbol{\delta T}$ are represented as a 'fixed-effect' term describing offsets between groups, $\boldsymbol{\alpha}$, and random effects describing temporal variations (five-year blocks), $\boldsymbol{\beta}_y$, and regional variations (17 sub-basin regions), $\boldsymbol{\beta}_r$. Matrices X, Z_y and Z_r specify, respectively, common pairs of groups, five-year blocks, and region. $\boldsymbol{\beta}_\sigma$ is the residual, and estimates are derived using an expectation-maximization procedure[36]. Groupwise SST corrections are applied in ICOADSb by removing estimated offsets from each SST measurement according to group, year and region.

Equation (1) is run at two levels, one for determining international offsets and one, a more detailed level, for determining interdeck offsets within nations. Each level of offsets is constrained to equal zero when summed across all paired measurements and all years. The groupwise corrections applied to ICOADSb thus adjust for offsets between groups but do not alter the average across all data. A detailed description of the LME design and implementation, along with the sensitivity of results to plausible variants, is available in a methods paper[10]. In an update to ref. [10], the analysis presented here uses international call signs for identifying nationality, thereby allowing us to increase the number of groups for which more than 5,000 SSTs are compared from 96 to 162 (Extended Data Fig. 10). Only groups associated with at least 5,000 pairs of SST observations are retained.

Trend estimates

Regional trends are the average of local trend estimates resolved at $5° \times 5°$ grid boxes at monthly resolution. Monthly errors are represented as the sum of four different components, $e \sim N\left(0, \sigma_o^2 + \sigma_s^2 + \sigma_b^2 + \sigma_g^2\right)$. Terms represent the uncertainty due to errors associated with individual observations, σ_o; partial sampling of each grid box, σ_s; HadSST3-type bucket-adjustment errors, σ_b; and errors common to individual groups of SST measurements, σ_g. The first three terms are assumed to follow those reported for HadSST3 (ref. [4]), and the last is inferred through the LME model.

A 1,000-member ensemble of SST observations is generated to represent both the random and the systematic components of uncertainty in ICOADSb. In addition to

uncertainties that are equivalent to those in HadSST3 (ref. [37]), groupwise bucket errors are included according to the results from our LME model. In particular, intergroup offset terms are drawn from a multivariate normal distribution that represents uncertainties associated with fixed group effects as well as random five-yearly and regional effects[10].

To compute trends, monthly anomalies are averaged to annual values, and years with fewer than three months of data are discarded. Empty monthly 5° grid boxes are infilled by averaging neighbouring grid boxes that are within 10°. Trends are reported only if SST estimates are present (or have been infilled) for both the first and the last five years, and data cover at least 26 of the years between 1908 and 1941. Trends are computed using standard linear least squares. The same procedure is followed for computing trends from the ensemble of realizations in order to estimate uncertainties.

The average trend uncertainty between 1908 and 1941 across sampled grid boxes is 0.12 °C per 34 years for $\sqrt{\sigma_o^2 + \sigma_s^2}$, 0.01 °C per 34 years for σ_b, and 0.06 °C per 34 years for σ_g. When taking spatial averages, σ_o and σ_s are independent across boxes, whereas σ_b is globally systematic and σ_g is partially systematic. The contributions of uncorrelated terms to uncertainties in the global mean trend are essentially negligible for σ_o and σ_s, remain at 0.01 °C per 34 years for σ_b, and become 0.05 °C per 34 years for σ_b. Groupwise errors are thus expected to dominate the uncertainties associated with SST trends.

Comparison with other data sets

There are a number of notable differences between the SST data sets considered here. ICOADSa, ICOADSb, and ERSST5 (ref. [1]) are based on ICOADS3.0 (ref. [22]), whereas COBESST2 (ref. [2]), HadISST2 (ref. [3]) and HadSST3 (ref. [4]) are based on ICOADS2.5 (ref. [38]. ICOADSa and ICOADSb use only bucket SST measurements, which are estimated to account for 94% of all observations in ICOADS3.0 between 1908 and 1941. Other data sets also make use of engine-room intake, buoy and drifter observations, which become more common after 1941. Finally, ERSST5, COBESST2 and HadISST2 infill monthly grid boxes without data, whereas ICOADSa, ICOADSb and HadSST3 leave these boxes unfilled.

Coastal near-surface air temperatures from CRUTEM4 (ref. [26]) are used to check the validity of our groupwise corrections. We choose to compare ICOADS with CRUTEM4 near the east coasts of Asia and North America because these two regions experience the largest adjustments in trends and because of the availability of relatively dense station and bucket data. In each area, regional SST and land air temperature time series are computed using only those grid boxes that contain both types of measurement (Extended Data Fig. 5c).

Pacific Decadal Oscillation

An SST index for the Pacific Decadal Oscillation, SST-PDO, is obtained by projecting annual-average SSTs poleward of 20° N in the Pacific onto a normalized PDO pattern, where the PDO pattern is obtained by regressing SST onto a standard National Center for Environmental Information (NCEI) PDO index[18] over the years 1948–2014 and is then normalized to have zero mean and a range of one over the North Pacific. The same method is applied to sea level pressure using the NOAA 20th Century Reanalysis[28], yielding SLP-PDO. SST-PDO is regressed against SLP-PDO over the periods 1908–1941 and 1948–2010 for sensitivities in units of °C mb^{-1}. Uncertainty in PDO trends is estimated by randomly perturbing PDO indices using error estimates of projections in individual years. When estimating the sensitivity of the SST-PDO index against the SLP-PDO index, the same random seeding is used to draw realizations of both indices for all SST products.
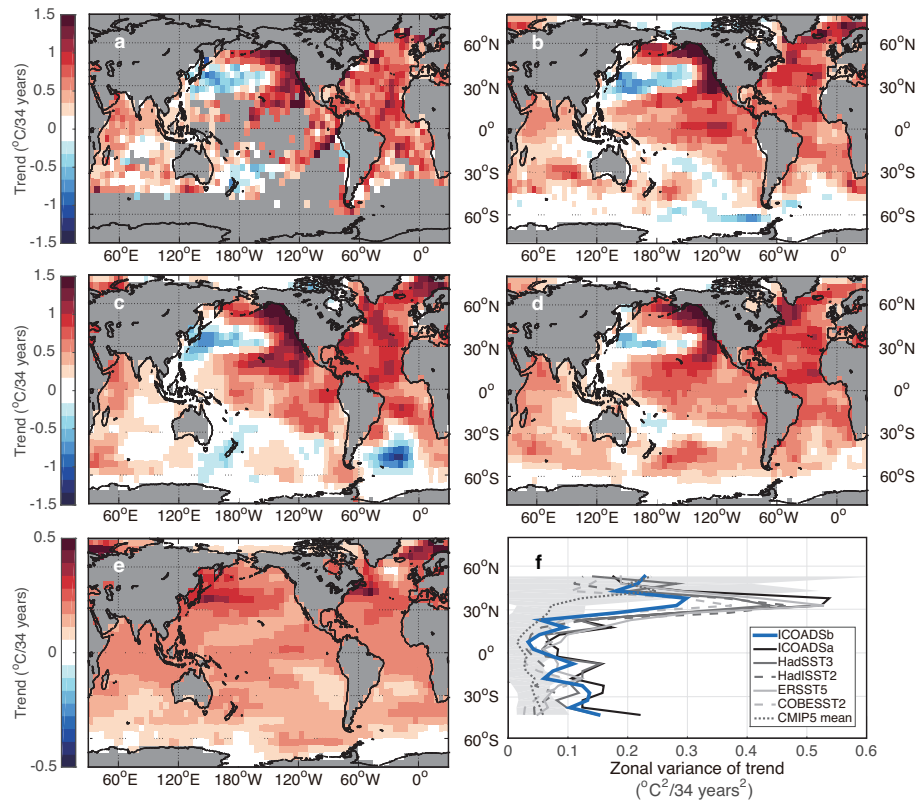
Data availability

All data sets used in this study are publicly available as follows: ICOADS3.0 (https://rda.ucar.edu/datasets/ds548.0/), HadSST3 and a 100-member ensemble (https://www.metoffice.gov.uk/hadobs/hadsst3/data/download.html), HadISST2 and a 10-member ensemble (https://www.metoffice.gov.uk/hadobs/hadisst/data/hadisst2/), COBESST2 (https://www.esrl.noaa.gov/psd/data/gridded/data.cobe2.html), and ERSST5 (https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.ersst.v5.html). ICOADSa and ICOADSb, together with archived versions of all other data sets, are posted at https://doi.org/10.7910/DVN/DXJIGA.
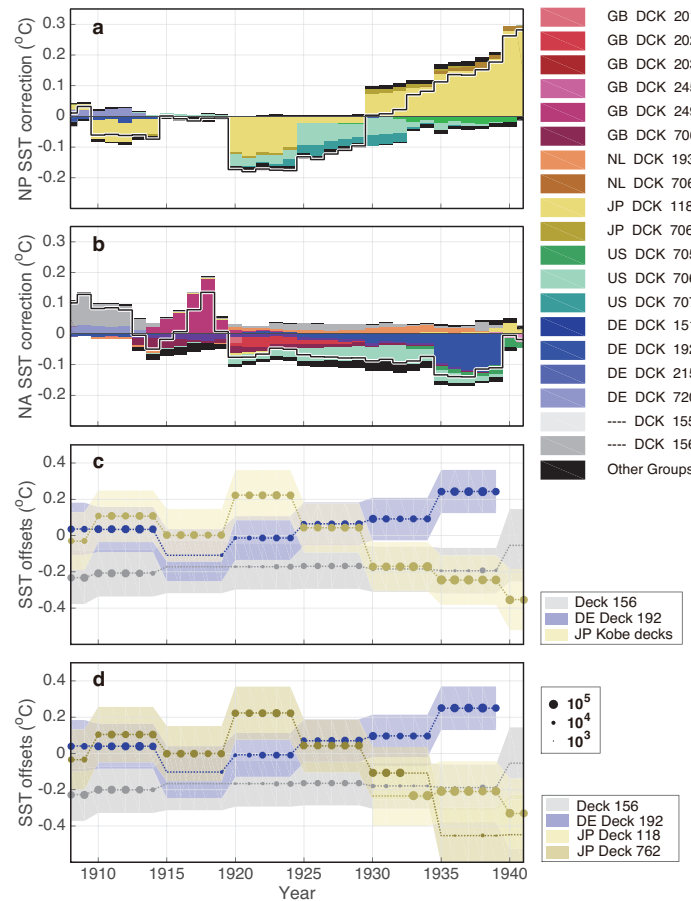
Code availability

Code allowing the full reproduction of our results is posted on Github at

https://github.com/duochanatharvard/Homogeneous_early_20th_century_warming.
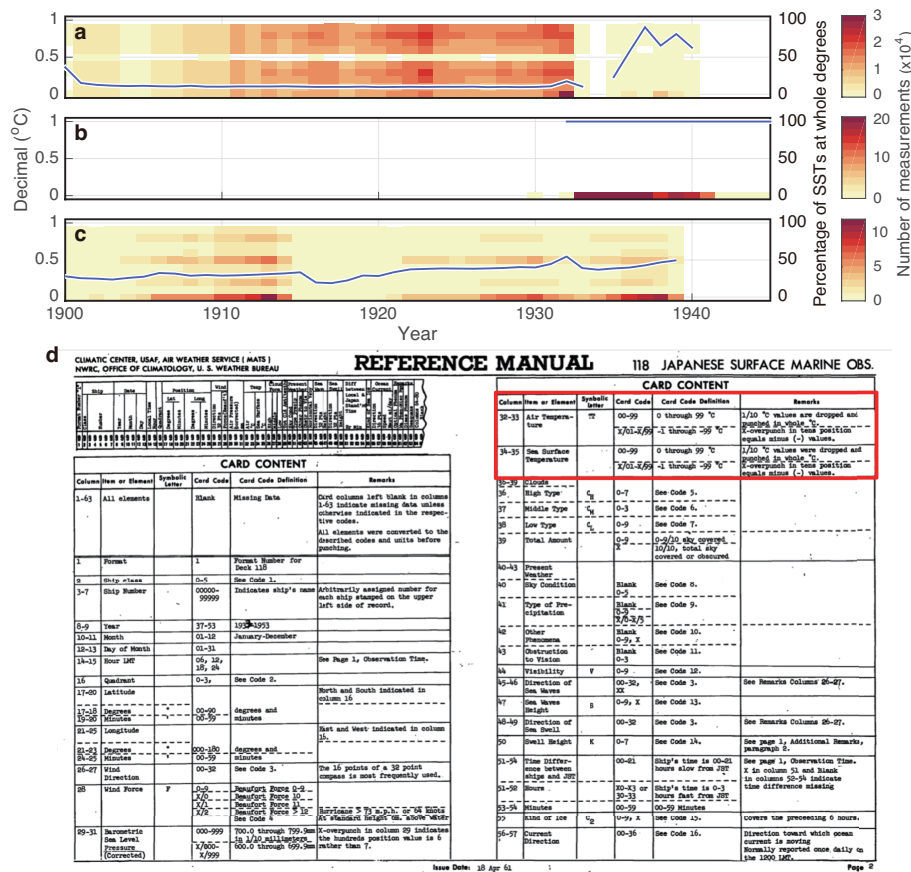
<jrn>31.Carella, G. et al. Estimating sea surface temperature measurement methods using characteristic differences in the diurnal cycle. *Geophys. Res. Lett.* **45**, 363–371 (2018).</jrn>

<bok>32.   *Radio Regulations*. Appendix 2, 793–799 (International Telecommunications Union, 2016).</bok>

<jrn>33.Carella, G., Kent, E. C. & Berry, D. I. A probabilistic approach to ship voyage reconstruction in ICOADS. *Int. J. Climatol.* **37**, 2233–2247 (2017).</jrn>

<jrn>34.Berry, D. I. & Kent, E. C. A new air–sea interaction gridded dataset from ICOADS with uncertainty estimates. *Bull. Am. Meteorol. Soc.* **90**, 645–656 (2009).</jrn>

<jrn>35.Dee, D. P. et al. The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).</jrn>

<jrn>36.Harville, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* **72**, 320–338 (1977).</jrn>

<jrn>37.Kennedy, J. et al. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res. D* **116**, (2011).</jrn>

<jrn>38Woodruff, S. D. et al. ICOADS release 2.5: extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* **31**, 951–967 (2011).</jrn>
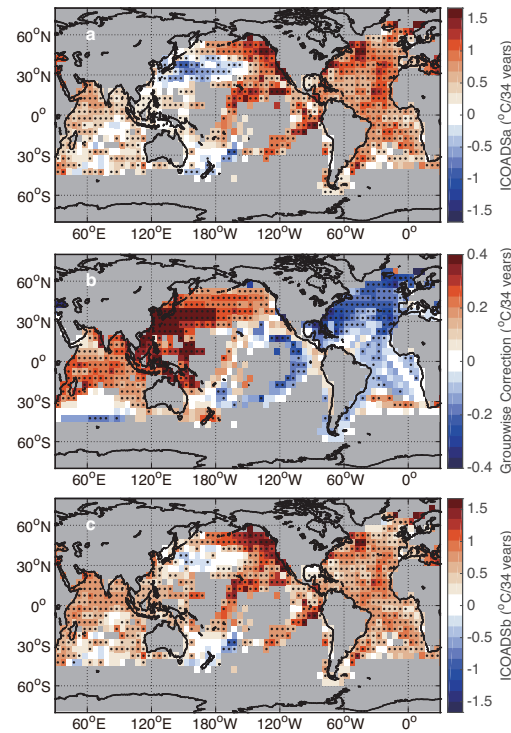
Extended Data Fig. 1 | 1908–1941 SST trends in major observational estimates and CMIP5 simulations. a–d, Maps of SST trends for: **a**, HadSST3 (ref. [4]); **b**, HadISST2 (ref. [3]); **c**, ERSST5 (ref. [1]); and **d**, COBESST2 (ref. [2]). Although different correction schemes were used to account for bucket biases in each observational estimate, panels **a–d** show similar early-twentieth-century trends. **e**, Averages from 88 CMIP5 historical experiment simulations[30]. Model results and observed trends are all regridded to a common 5°×5° resolution to facilitate intercomparison. **f**, Zonal variance in 1908–1941 SST trends. Shading indicates a range covering 80% of the 88 CMIP5 members, which come from 39 coupled climate models (see Supplementary Table 4). ICOADSb has a more uniform spatial pattern of early-twentieth-century warming than other SST estimates.
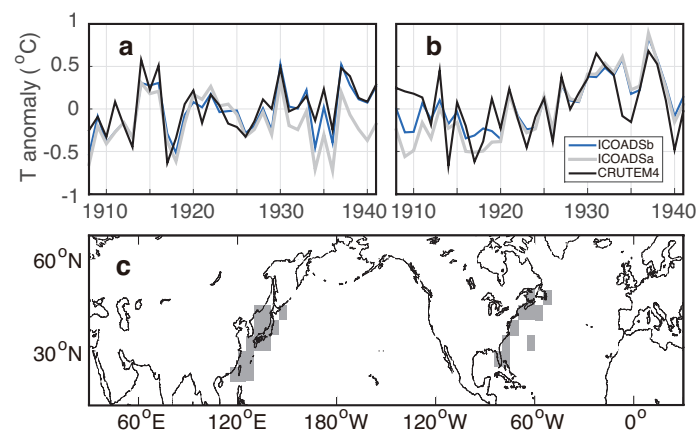
Extended Data Fig. 2 | Groupwise decomposition of SST corrections in ICOADSb. a, **b**, Contributions from individual groups over the North Pacific Ocean (NP; **a**) and the North Atlantic Ocean (NA; **b**). **c**, SST offsets for groups having major influences on 1908–1941 trend estimates. Groupwise corrections relate foremost to a growing cold bias amongst Japanese deck 118 (yellow) in the Pacific, a pre-World War I cold bias in deck 156 (grey), and a growing warm bias in Germany deck 192 (blue) in the Atlantic. Shading indicates 2 s.d. uncertainties, and the sizes of the markers indicate numbers of SST observations. **d**, As for **c**, but for an analysis in which the Japanese Kobe collection is divided into decks 762, 118 and 119. In this analysis, the mean offset over the North Pacific is $0.07 \pm 0.14\,°C$ (2 s.d.) for deck 762 from 1908 to 1930, and $-0.24 \pm 0.16\,°C$ for deck 118 from 1935 onward, consistent with the case when decks are not separated (**c**). In addition, the cold offset of the Japanese Kobe Collection and the warm offset of German deck 192 are robust even if SSTs from the two groups are not allowed to be directly paired in the intercomparison.

Extended Data Fig. 3 | Decimal distributions of SSTs. a, Deck 762 from the Japanese Kobe Collection. The near-absence of a 0.5 °C decimal is indicative of data being recorded in whole-degrees Fahrenheit and then converted into tenths-of-a-degree Celsius. The blue line shows the percentage of SSTs recorded in whole degrees. **b**, Temperatures reported in deck 118 of the Japanese Kobe Collection are all truncated to whole-degrees Celsius and constitute more than 99% of the Kobe Collection measurements between 1933 and 1941. **c**, Germany deck 192 is a time-varying mix of data in whole-degrees Celsius and data in tenths-of-a-degree Celsius; most common in the latter are values of 0 and 5, and then even digits. The percentage of SSTs archived in whole-degrees Celsius increases from 29% in 1908 to 50% in 1941 (blue line). **d**, Image from the reference manual of Japanese Kobe Collection deck 118 (ref. [25]), where the red box highlights that temperature measurements were digitized at whole-degrees Celsius and all decimals dropped (https://icoads.noaa.gov/reclaim/pdf/dck118.pdf).
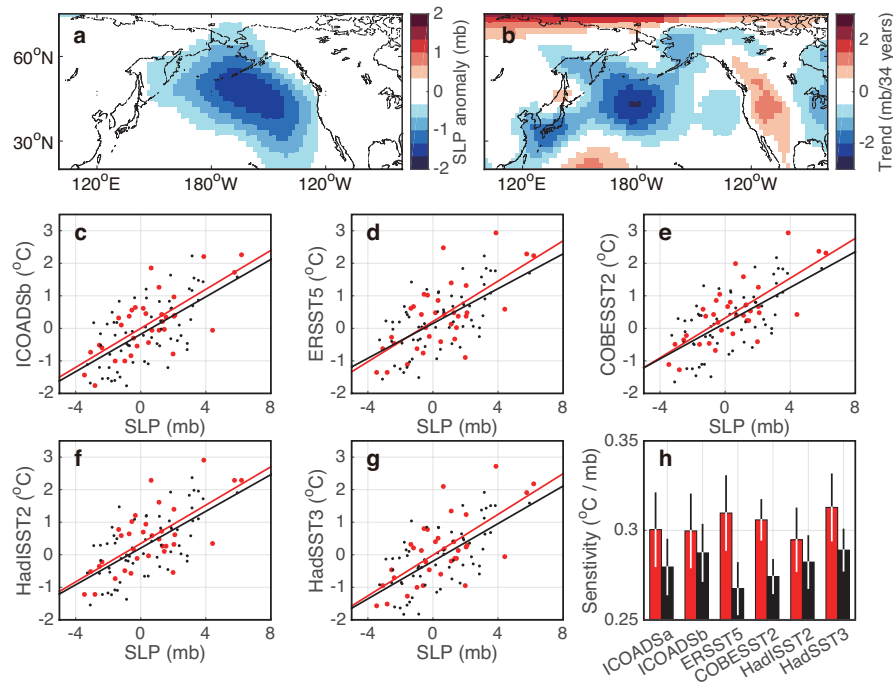
Extended Data Fig. 4 | Sensitivity of 1908–1941 SST trends. As for Fig. 3, but with Japanese Kobe Collection decks 118, 119 and 762 separated into distinct groups. **a**, SST trends in ICOADSa; **b**, trends associated with the corrections for groupwise offsets; **c**, SST trends in ICOADSb after applying groupwise corrections.
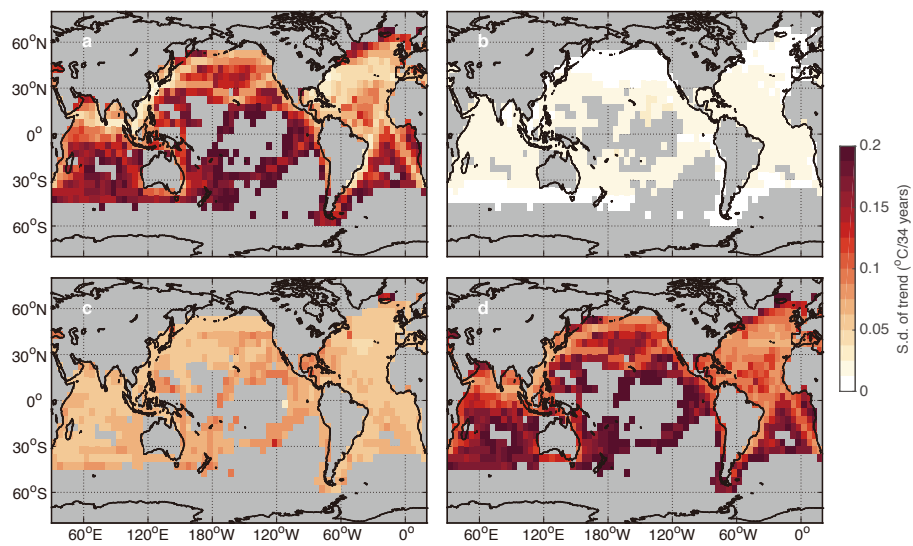

Extended Data Fig. 5 | Comparison of SSTs with coastal air temperature estimates. a, b, Air temperatures are from CRUTEM4 (ref. [26]) near the east coast of Asia (**a**) and the east coast of the US (**b**). These two regions are indicated by the shading in **c** and host the largest groupwise corrections. SSTs corrected for groupwise offsets (ICOADSb, blue line) better
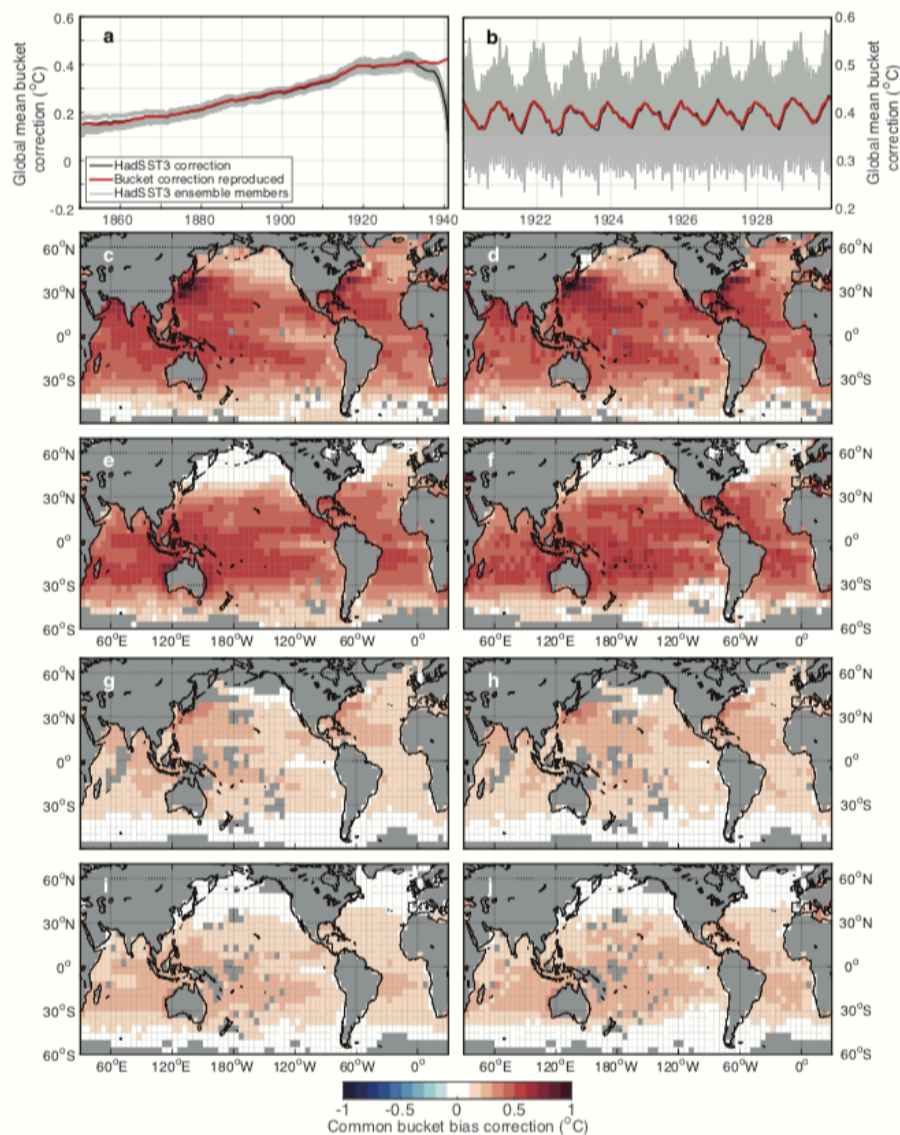
correlate with CRUTEM4 air temperatures (black line) than do SSTs having only common

bucket-bias corrections (ICOADSa, grey line), especially for Asia in the 1930s. See Table 1 for

correlations associated with other major SST estimates.

Extended Data Fig. 6 | Sea level pressure and the PDO. a, Spatial pattern of sea level

pressure (SLP) associated with a 1σ increase in the PDO index[18]. **b**, 1908–1941 SLP trends

from the NOAA's 20th Century Reanalysis[28]. **c–g**, Regressions of SST-PDO indices against

the SLP-PDO index for ICOADSb (**c**), ERSST5 (**d**), COBESST2 (**e**), HadISST2 (**f**) and

HadSST3 (**g**). The SLP-PDO index is computed from the NOAA's 20th Century Reanalysis,

and the regression analysis is performed for 1908–1941 (red) and 1948–2010 (black). **h**,

Regression coefficients for each data set over 1908–1941 (red) and 1948–2010 (black). Note

that the y-axis starts from 0.25 °C mb^{-1}. Error bars indicate 2 s.d. uncertainties. ICOADSb has

a similar regression coefficient over 1908–1941 to that found using other SST data sets and
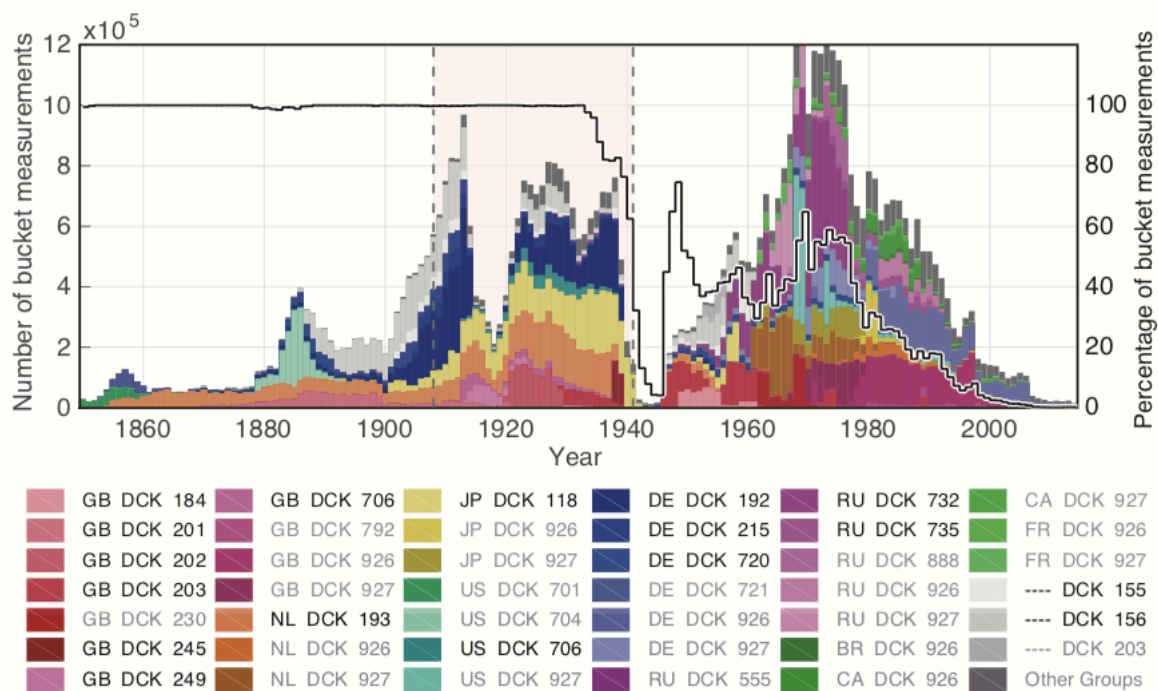
other epochs.

Extended Data Fig. 7 | 1σ uncertainty estimates of 1908–1941 trends. a, Sampling and observational uncertainty inherent to all SST estimates, estimated using 1,000 members drawn from normal distributions that have a mean of zero and a variance of the sampling and observational uncertainty equal to that specified in ref. [37]. **b**, Uncertainties associated with common bucket-bias corrections in HadSST3, which takes into account shifts between wooden and canvas buckets, exposure time, and ship speed, obtained from a 100-member ensemble[4]. **c**, Uncertainty of groupwise corrections in ICOADSb, estimated from 1,000 random correction members. **d**, Uncertainties quadratically combining those shown in **a**–**c**. Note that these uncertainties are for individual grid boxes, whereas Table 1 reports uncertainties for regional or global averages.
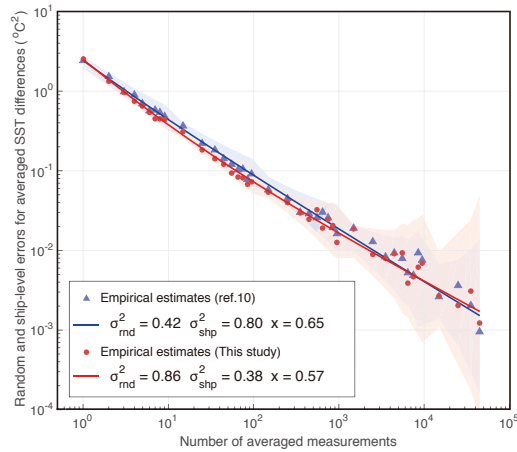
Extended Data Fig. 8 | Common bucket-bias corrections. a, Comparison of annual global mean common bucket-bias corrections reproduced in ICOADSa and from HadSST3 for 1850–1941. **b**, As in **a**, but for monthly mean corrections from 1920 to 1929. Bucket corrections in ICOADSa are subsampled over HadSST3 grids before computing global means. **c–f**, Comparisons of spatial patterns of SST corrections for the period 1920 to 1930. Shown are our corrections for December, January and February (**c**) and June, July and August (**e**); and the HadSST3 corrections for December, January and February (**d**) and June July and August (**f**). The 1920-1930 period is assumed to contain only canvas buckets, and an average of 78% of SSTs from fast ships (7 m s^{-1}) and 22% from slow ships (4 m s^{-1}). **g–j**, As for **c–f**, but for

corrections over 1850 to 1860—a period assumed to contain only slow ships (4 m s^{-1}), with 36% of SST measurements from canvas buckets and 64% from wooden buckets. Discrepancies become apparent from the 1930s because of increasing contributions from warm-biased engine-room-intake measurements in HadSST3, whereas we focus only on bucket measurements.

Extended Data Fig. 9 | Groups providing bucket SST measurements from 1850 to 2014. Left-hand y-axis: numbers of bucket SST measurements from individual groups, identified by country and deck information as colour coded, in ICOADS3.0. CA, Canada; BR, Brazil; DE, Germany; FR, France; GB, Great Britain; NL, Netherlands; JP, Japan; RU, Russia; US, United States; —, missing country information. Black text shows the 46 of 162 groups that contributed to the SST archive between 1908 and 1941. Groups with fewer than 2×10^5 measurements throughout 1850 to 2014 are labelled as 'other groups'. Right-hand y-axis, percentage of measurements that are from buckets (black line). Bucket measurements constitute nearly all observations before the 1930s.

Extended Data Fig. 10 | Partitioning of random and systematic errors. Updating the methodology for identifying nationality allows more groups of SSTs to be compared, and leads to minor changes in estimated random and systematic errors. Shown are fits according to equation (6) in ref. [10], depending on random measurement error (σ_{rnd}^2), systematic ship-level error (σ_{shp}^2), and the scaling relationship between the number of measurements and ships (x). Fits are performed using logarithms, and shading indicates 2 s.d. in logarithmic space.